LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-428046

An Autonomous System for Grouping Events in a Developing Aftershock Sequence

D. B. Harris, D. A. Dodge

April 20, 2010

Bulletin of the Seismological Society of America

Disclaimer

An Autonomous System for Grouping Events in a Developing Aftershock Sequence

D. B. Harris[1,2], D. A. Dodge[1]

[1]Lawrence Livermore National Laboratory

[2]Deschutes Signal Processing LLC

April, 2010

Abstract

We describe a prototype detection framework that automatically clusters events in real time from a rapidly unfolding aftershock sequence. We use the fact that many aftershocks are repetitive, producing similar waveforms. By clustering events based on correlation measures of waveform similarity, the number of independent event instances that must be examined in detail by analysts may be reduced. Our system processes array data and acquires waveform templates with an STA/LTA detector operating on a beam directed at the P phases of the aftershock sequence. The templates are used to create correlation-type (subspace) detectors that sweep the subsequent data stream for occurrences of the same waveform pattern. Events are clustered by association with a particular detector. Hundreds of subspace detectors can run in this framework a hundred times faster than real time. Nonetheless, to check the growth in the number of detectors, the framework pauses periodically and re-clusters detections to reduce the number of event groups. These groups define new subspace detectors which replace the older generation of detectors. Because low-magnitude occurrences of a particular signal template may be missed by the STA/LTA detector, we advocate restarting the framework from the beginning of the sequence periodically to reprocess the entire data stream with the existing detectors.

We tested the framework on 10 days of data from the NVAR array covering the 2003 San Simeon earthquake. 184 automatically-generated detectors produced 676 detections resulting in a potential reduction in analyst workload of up to 73%.

Introduction

The resources of network operations often are severely taxed during the aftershock sequences that follow major seismic events. Many thousands of aftershocks often occur in a period of weeks following a major event, which may overwhelm analysts tasked to review each individual event. However, some scope exists to increase the efficiency of event review due to the fact that a significant fraction of aftershocks occurs in clusters and, consequently, produce similar waveforms. Aftershocks often appear to be concentrated on the margins of asperities [Lay and Wallace, 1995], so that the source region may come to resemble a field of discrete nucleation sites with some background of diffuse seismicity. If a reliable procedure can be devised automatically to cluster related aftershock waveforms prior to analyst review, it may be possible to present waveforms in related families. Detailed analysis may be required only on a single event within each family.

To be useful in operations, such a system must discover repeating event patterns and assemble related event waveforms autonomously as aftershock sequences unfold. The very substantial body of work on event cluster detection by waveform correlation [e.g. Israelsson, 1990; Nadeau et al., 1995; Nadeau and McEvilly 1997; Schaff and Richards, 2004; Schaff and Waldhauser, 2005; Schaff, 2009] can be drawn upon for automated discovery of repeating event signatures. Some means of making initial detections may be employed to define an event pool, from which repeating signatures are discovered through clustering on waveform correlation measurements. Correlation and subspace detectors [e.g. Gibbons and Ringdal, 2006; Harris, 2006; Schaff, 2009] then provide the means to sweep the subsequent data stream for occurrences of defined patterns (including low-magnitude events not detectable by

power detectors). Combining waveform correlation clustering and correlation detection at the front end of a processing pipeline may support efficient analyst review.

If the association of events by waveform correlation can be trusted, for example by use of a suitably high correlation threshold, then a reasonable measure of workload reduction may be based upon analysis of a single event within each cluster in place of all events. For example, if 1000 events are detected, but fall into 300 clusters, then a 70% reduction in workload may be achieved.

While attractive in theory, the notion of automatically detecting waveform patterns and spawning correlation detectors in real time to screen them has practical pitfalls. First, in any automatic system, there could be a tendency to generate detectors from noise bursts and other undesired transients that would cause the number of correlation detectors to proliferate undesirably. Second, in any aftershock sequence, many events will occur with superimposed waveforms complicating the task of choosing a data window to define the waveform pattern or template to represent a repeating source. In our experience this latter problem can be very subtle and lead to multiple detectors searching for a single pattern. An undesirable result can be large numbers of duplicate detections of events in several automatically-defined clusters. Any practical system must implement procedures to minimize these problems.

Even a properly functioning system designed to minimize duplication can have a large number of detectors. Correlation detectors have a limited source-region footprint (typically about one wavelength at the dominant frequency of the template waveform), so that large numbers of them are required to cover a source region that can span thousands of square

kilometers. Consequently, careful attention to the details of signal processing is warranted to increase the speed of correlation detectors. In addition, policies to limit the number of detectors need to be implemented. We elected to use generalizations of correlation detectors (subspace detectors) in an attempt to reduce the number of screening detectors. We introduced a supervisory function to the system that periodically compares detectors and eliminates or combines duplicates. We also developed implementations of subspace detectors that run about 10,000 times faster than real time on commodity microprocessors. Consequently, the hundreds of detectors created by an autonomous system has not proven to be a computational limitation.

This paper is organized in four additional sections. We describe methods in the next section, specifically the prototype autonomous framework that we implemented to create, spawn and, periodically, consolidate subspace detectors on a developing aftershock sequence. The following section briefly presents the data we used to test the system: 10 days of continuous data from the NVAR array covering the initial aftershock sequence of the 2003 San Simeon earthquake. The third section presents results, which show a significant potential for increasing analyst efficiency. The final section presents our conclusions, in which we summarize what our experience indicates for the larger implications of repeating seismicity for network operations, limitations of this study, and the potential for generalization to correlation detection and screening across networks (as opposed to an individual array).

Methods

Subspace Detectors

We chose to use subspace detectors as our principal method to search data streams for repeating events. Subspace detectors [Scharf and Friedlander, 1994; Harris, 2006; Harris and Paik, 2006; Maceira, et al., 2010] are generalizations of waveform correlation detectors [see e.g. Gibbons and Ringdal, 2006; Gibbons et al., 2007; Schaff, 2008] that permit a degree of variation in the signals being detected. Subspace detectors are attractive for seismological applications because seismic signals rarely repeat exactly. The variation of related waveforms can be represented in a waveform basis constructed from a sample ensemble of such waveforms. The basis is constructed by assembling ensemble waveforms in sampled (digital) form as columns of a data matrix, then performing a singular value decomposition (SVD) of the data matrix to extract a low-rank orthonormal set of spanning vectors (left singular vectors). Prior to assembling the data matrix, the waveforms are aligned using shifts obtained by maximizing the cross-correlation functions of the event waveforms [VanDecar and Crosson, 1990]. Typically the rank (i.e. dimension) of the subspace representation so obtained is determined from the number of appreciable singular values. We estimate the rank as the number of singular values needed to represent a given fraction (typically 90%) of the energy in the eigenspectrum of the data matrix. Correlation detectors are rank-one subspace detectors; i.e. the signal subspace is characterized by a single basis waveform.

Subspace detectors operate by sliding a detection window over the data stream, extracting the (multichannel) waveform in the detection window and approximating the extracted waveform as the best (in the least-squares sense) linear combination of the basis

waveforms. The detection statistic is formed as the energy in the approximating waveform divided by the total energy in the extracted waveform. The normalization makes the detector a constant false alarm rate (CFAR) detector, which means that its threshold can be set independently of background noise level. The detection statistic ranges between zero and one, and is equal to one if and only if the waveform in the detection window is exactly a linear combination of basis waveforms. The detection statistic can be determined alternatively as the square of a running correlation between the observed waveform in the detection window and its best linear approximation from the subspace basis. In the rank one case, the statistic reduces to the square of the familiar correlation coefficient between the observed signal and the desired waveform template.

Significantly more detail about the design and implementation of subspace detectors can be found in Harris [2006] and [Harris and Paik [2006], available as electronic supplements to this paper.

Detection Framework

Figure 1 provides a high-level overview of the framework we developed to test the idea of an autonomous, real-time system for creating subspace detectors. The heart of the system is a dynamically-configurable list of detectors that can hold an arbitrary number of instances of several types of detector. These detectors are applied to a common stream of array data, processing the data in consecutive, contiguous blocks. For each block, the framework directs each detector in the list to calculate a detection statistic and examine that statistic for excursions above a predetermined threshold. Detectors produce triggers whenever the threshold is exceeded, and often more than one detector will trigger on the same event. The

goal is to declare a unique detection for each event, associated with one detector. Consequently, it is necessary to have a policy for comparing triggers and promote one to the status of detection. We describe our policy below. In our test application to the San Simeon sequence, the detector list holds a single, permanent, power (STA/LTA) detector operating on an array beam configured for San Simeon region Pn and an arbitrary number of subspace detectors, dynamically added and removed. The function of the power detector is to supply new signal templates to the system which uses them to create and add new subspace detectors to the list.

Since the purpose of the system is to associate events based on waveform correlation, our policy for resolving conflicting triggers favors subspace detectors over the beamforming power detector. We have two rules:

(1) Triggers from the same type of detector are promoted or eliminated based on which has the largest detection statistic value. The trigger with the largest statistic is promoted, the rest are eliminated.

(2) Triggers from subspace detectors always are promoted over those from power detectors.

All detections are archived with information about the detector that originated them, the trigger time and the triggering value of the detection statistic. Detections from the power detectors are assumed to be signals not yet seen: the system passes such detected waveforms through a series of screens (for example on duration and bandwidth) in an attempt to eliminate spikes and other unwanted signal types. Experience with overlapped events led us to

implement a check of a proposed correlation template against the templates of all existing detectors. If a proposed template is found to have a significant projection onto an existing correlation detector template, the proposed template is rejected. Waveforms that pass all screens are used to create rank-1 subspace (correlation) detectors which are added to the detector list. The ownership of the detection used to create a new detector passes from the power detector to the newly-created subspace detector. As described earlier, the system processes the data in sequential, contiguous blocks. The operation of adding new detectors to the detector list occurs at the conclusion of processing a block, just before the framework moves on to the next block.

Periodically the system halts to consolidate subspace detectors in the list, and then continues processing the data stream. The purpose of consolidation is to provide some check on the otherwise unrestricted growth in the number of subspace detectors. During consolidation, detections from all recently-constructed subspace detectors and previously-ungrouped detections are put through a clustering operation to define new templates. Consolidation is initiated when the number of events detected since the previous consolidation episode or the beginning of operation exceeds a threshold (300 in our test). The events subject to consolidation are extracted from the archive (detection pool in Figure 1) and correlations are calculated between the waveforms for all distinct event pairs. A single-link algorithm is used to cluster the events based upon the correlation values and the correlation lags are used to align event waveforms within each cluster. A subspace detector template is constructed from the aligned waveforms from each cluster, and the collection of subspace detectors so created is added to the detector list. Detectors are removed from the list based upon a concept of

transfer of detection ownership. The ownership of events in a cluster produced during consolidation is transferred from the originating detectors to the subspace detector newly created from the cluster. If transfer of ownership leaves any detector without detections, that detector is eliminated from the list.

This rather complicated policy for consolidating detectors was the end point in an evolution of techniques we tried to check the growth in numbers of detectors. Throughout this evolution, we were concerned that it would take some time for enough aftershocks to accumulate for major repeating patterns to become apparent. Retiring some detectors too early might only result in the necessity to reestablish detectors for the same signal pattern later. In the interim, detections of smaller events fitting the pattern might be missed.

At first we implemented a policy that subspace detectors, once spawned from their initial power detection template, were permanent. However, we found that many events produced triggers by more than one detector created this way. Consequently, the waveform patterns defining these detectors have some projection onto each other. We implemented a policy of tracking triggers and detections associated with each detector to find detector pairs that had a large number of triggers in common. We intended then to replace such detector pairs by pooling their associated events and using the events to define a single, probably higher-rank subspace detector. What we found from this exercise was that often many more than two subspace detectors were related by common triggers. It proved to be too complicated to choose pairs of detectors for consolidation.

As a consequence, we contemplated a policy of pooling detections from all existing detectors, producing clusters from that pool and generating new detectors from the clusters, which would replace the existing suite of subspace detectors. However, the computational burden of the waveform correlation calculation grows as the square of the number of events, and clearly would be infeasible after a long period of operation. We decided to divide the subspace detectors into two groups to limit this calculation. We designated subspace detectors as either generation 1 or generation 2 detectors. Generation 1 detectors are those created directly from a beam power detection (and are always rank 1 detectors). Generation 2 detectors are those created from a consolidation operation and can be higher rank. We assumed that enough events accumulate by the time that consolidation occurs that patterns are well established and the generation 2 detectors need not be replaced in subsequent processing. This conjecture allowed us to exclude detections produced by generation 2 detectors from the pool. This fact along with the natural slowing in the rate of aftershock occurrence with time (according to Omori's law) acceptably bounds the correlation calculation. The net result is that the correlation pool consists of the detections made since the most recent consolidation round augmented by any stragglers that failed to be grouped in prior rounds.

Data and Processing Parameters

For a suitable test, we required an event with large numbers of aftershocks in a short period of time and good ground truth information, observed by an array at regional distance with many high-SNR observations suitable for generating correlation templates. We chose the 2003 San Simeon earthquake (Figure 2). This was a moderately large event (mb 6.5) with

thousands aftershocks recorded by local networks in California [Hauksson et al., 2004; McLaren et al., 2008]. For ground truth information on the aftershocks, we relied upon the Advanced National Seismic System (ANSS) composite catalog [NCEDC, 2009] which reported 1433 events in a 1x1 degree square around the main shock during the last ten days of 2003. We acquired ten days of data (2003:356 – 2003:365) from the NVAR array for the test (see Data and Resources section). NVAR is an IMS primary station, an array configured (Figure 2 inset) for observing regional phases, approximately 390 kilometers from the main shock location. We chose to use 9 channels of the array due to data quality problems at the other array elements.

Template window selection

The San Simeon aftershock waveforms observed by the NVAR have durations on the order of 100 seconds (Figure 3). We chose to use a 110 second data window to define signal templates, beginning approximately 10 seconds before the P onset. In selecting template windows, we used an estimate of the P onset determined from the point where the power (STA/LTA) statistic exceeded its pre-defined detection threshold. We filtered the data into a 1-3 Hz passband, computed a beam directed at the Pn phase for the San Simeon region, squared the signal and used a running STA of 3 seconds duration and LTA of 30 seconds duration to detect the P phase. The threshold on the STA/LTA ratio that we used was set to a relatively large value (25 in power, corresponding to 5 in signal amplitude) in order to obtain high-SNR waveform templates with fairly accurate estimates of the P onset.

As in all aftershock sequences, it frequently happens that such a large template window will encompass the arrival of signals from two events. Figure 4 shows an example of this type of occurrence, where the STA/LTA algorithm triggered on the P arrival from a small event (A),

which was then followed approximately 55 seconds later by the signal from a much larger event (B). This particular detection caused significant problems in the detection framework, because the signal from event B, being the more energetic, came to dominate detections by the template formed from this data window. It happened that an STA/LTA trigger on B had already caused the formation of another subspace detector aligned to the onset of B. These two detectors then produced duplicate detections on a large set of aftershocks. The duplicate detections could not be eliminated by the usual comparison of simultaneous triggers, because the triggers on the template formed from the event A detection were labeled with a time 55 seconds earlier than the triggers on the template formed from the event B detection. Our test for simultaneity required triggers to occur within 20 seconds of each other (even this is a fairly liberal coincidence threshold). Our solution to this problem of duplicate detections was to modify the trigger comparison code to use a coincidence threshold of 100 seconds for STA/LTA triggers, and 10 seconds for subspace triggers.

Subspace Detection Threshold

The key parameter influencing the number of events detected and clustered by the detection framework is the threshold for correlation and subspace detection. We set this value to 0.1, which corresponds to a standard correlation coefficient of 0.316. This value is lower than commonly is used in many studies involving waveform correlation and requires some justification. Our motivation in setting a low threshold was to increase the size of clusters in support of efficient event review. At the same time, the results of clustering have to be trusted,

meaning that the events within a cluster are related, so that an interpretation developed for one can reasonably be applied to the rest.

High, chance correlation values are most likely to occur among unrelated events at similar distances from the observing station, since all the phases in the corresponding seismograms will be approximately aligned. To assure meaningful clusters, we chose to set the correlation threshold significantly higher than an empirical distribution of correlation values formed from unrelated events at approximately the same distance from the NVAR array as the San Simeon sequence. To calculate this distribution, we assembled 14 events in the same distance range as the San Simeon sequence (Figure 5). The waveforms of these events recorded on one channel of the array are displayed in Figure 6, which shows that the inter-phase arrival times are similar, as expected. The distribution of correlations for distinct pairs of signals in this set is shown in Figure 7. For comparison, we also plot the selected correlation threshold as a vertical bar. The correlation values are confined below 0.03, significantly less than the threshold. The distribution of sample correlations for this event set is so close to zero because we use a large amount of data to define the waveform correlation template (9 channels x 110 seconds per channel x 2 Hz bandwidth). The theoretical time-bandwidth product of the observations is very large (~4000), which makes chance high correlations very unlikely. We will discuss additional evidence that clusters at this threshold value are meaningful in the next section.

San Simeon Sequence Results

To establish a baseline for assessing the results of the detection framework, we processed the data with the beamforming power detector alone. The parameters were as described earlier: bandwidth 1-3 Hz, STA duration 3 seconds, LTA duration 30 seconds, power threshold 25, beam directed at Pn velocity for the great circle backazimuth to the San Simeon source region. We also implemented a post-detection blackout window of duration 30 seconds. This processor produced 306 detections in the ten day period, of which 138 (45%) were associated with the San Simeon sequence by reconciliation (automatic with manual review) with the ANSS catalog.

The test of the framework consisted of two passes over the data, the first with the spawning and consolidation features turned on to create subspace detectors, and the second with those features turned off. However, on the second pass, the collection of subspace detectors created by the end of the first pass was held fixed and used to scan the entire 10 day data record. The purpose of the second pass was to use detectors developed late in the sequence to scan for events of the same waveform pattern, but lower SNR, that occurred earlier in the sequence. Since an SNR threshold was imposed on power detections used to spawn correlation detectors, early small events might be lost until a suitable larger template event occurred. The first pass produced 184 detectors and 496 detections . The second pass produced 676 detections with the same 184 detectors; 387 (57%) of these detections were reconciled with San Simeon events in the ANSS catalog.

Figures 8 and 9 summarize the results of the beamforming detection test and the two-pass detection framework test. Figure 8 displays the geographic distribution of events in the

source region as grey symbols for the events reported in the ANSS catalog, but not detected by either processor, as black symbols for the events detected by both the beamforming algorithm and the detection framework, and as red symbols for the events detected by the detection framework alone. In a sense, the red events represent the value added by the detection framework over the beamforming detector. Figure 9 presents the same data as histograms of event occurrence or detection as a function of magnitude. From this latter plot it is apparent that the detection framework principally adds event detections at magnitudes below 3 as would be expected by the fact that it employs correlation-type detectors. However, it does pick up a significant number of events missed by the beamformer above magnitude 3 and even a few above magnitude 4. The likely cause of this improved performance at higher magnitudes is the fact that correlation and subspace detectors compress all of the energy in an event waveform into a single spike, which enables detection of smaller events in the coda of larger events. In our system, the post-detection blackout period for subspace detectors was 3 seconds.

The reduction in interpretation workload might be reckoned in one of two ways from these tests. If we assume that the additional detections produced by the system add no value (probably an unreasonable assumption), then the number of events requiring interpretation has been reduced from 306 to 184. The latter number is the number of groups of detections associated with the 184 detectors created by the framework. In this interpretation, the workload has been reduced by 40%. If instead, the 676 detections are considered to be valuable to the network operation, then the workload is reduced by 73%, i.e. by almost a factor of four.

Examples of the clusters of event waveforms formed automatically by the detection framework are shown in Figures 10 and 11. These contain events associated with three detectors. Within each group, and despite the relatively low value of the correlation threshold, there are broad similarities in waveform shape, though also significant variations. However, the waveform characteristics differ more strongly between groups than within each group. Nonetheless, the fact that significant waveform variations occur with each group has prompted us to take care with the argument that the clustered events are indeed related and can be interpreted as a group rather than individually.

Perhaps the strongest line of evidence that the event groups represent related events is the fact that they tend to form compact geographic clusters. Figure 12 shows the events associated with 6 different detectors (listed by an automatically-generated number), where color now encodes membership in a particular detector event cluster. The geographic footprints of the event groups generally are small, with a few outliers that may be due to errors in reconciliation against the ANSS catalog. Our reconciliation algorithm was based on the coincidence of an origin time estimated from the detector trigger time and the catalog origin time. Frequently, many events occurred close together in time. In addition, the largest group of detections was associated with detector 116, a rank-2 subspace detector. The fact that it has more degrees of freedom in its waveform representation may account for the presence of legitimate geographic outliers.

A significant concern with any system based upon automatic generation of correlation detectors is that “nuisance” detectors may be generated that clutter the system with many

detections of small events or similar noise bursts. A system employing automatic detector generation will not be a net improvement over standard pipelines if too many detections of small, insignificant signals are produced. Our experience is, that, although unintended detectors are created by the system and many additional detections are produced, these are either legitimate local events (Figure 13) or perhaps low time-bandwidth product noise bursts (Figure 14), that are simple and quick to note and dismiss when organized into groups. They are subject to the same economy of interpretation that we claim for the intended target sequence.

Conclusions

Based on this preliminary work, it appears likely that network pipeline operations could be made significantly more efficient during aftershock sequences by incorporating an automatic capability to generate detectors designed to detect and group similar events. Our experience further indicates that higher efficiencies can be obtained by reprocessing the data from the beginning of the sequence periodically with maturing suites of detectors. We speculate that even routine operations may benefit from a detector spawning framework in situations where network stations are located in mining regions or areas of high natural seismicity.

Periodic reprocessing is feasible because the number of detectors can be managed through a supervisory function to replace proliferating single-rank detectors with more general single- or higher subspace detectors and because these detectors can be implemented O(10,000) times faster than real time. Even with our research code heavily instrumented for debugging purposes, processing 10 days of data with several hundred detectors (and

performing consolidation operations periodically) typically took 30-35 minutes on a single commodity processor. An operational system could be parallelized with detectors operating in different threads perhaps implemented on GPU cards to obtain far greater speeds.

Consistent with Omori's Law of exponential decay in the rate of aftershock occurrence, one can imagine a system restarting at log periodic intervals, perhaps at two hours, then four, then eight and so on to provide analysts with the most complete set of event clusters possible with recently obtained information. Base on our preliminary experience, it appears computationally feasible to keep this type of operation running for weeks or a few months until an aftershock sequence subsides. The system, once restarted at the outset of the sequence would quickly catch up to real time. Indeed, it may be desirable to keep two instances of the system running simultaneously: one reprocessing the sequence and the other processing forward in real time.

The system we implemented and tested was a partial realization of the kind of detection framework that would be required for network operations. A more complete test would involve spawning subspace detectors in the context of a network of arrays operating suites of beamforming detectors (beam recipes). A system designed to operate on each array independently would be a straightforward generalization of the prototype described here, however coordination across stations in a network to develop correlation-type detectors processing coherently across a network is a more difficult problem.

Data and Resources

The NVAR array waveform data used in this study were obtained from the U. S. National Data Center at the Air Force Technical Applications Center, Patrick Air Force Base, Florida. These data can be obtained from the IRIS Data Management Center at www.iris.edu. The catalog data used as ground truth information was obtained from the Advanced National Seismic System Composite Earthquake Catalog at www.ncedc.org/cnss/. Maps were produced using the Generic Mapping Tools version 4.1.3 (www.soest.hawaii.edu/gmt; Wessel and Smith, 1998).

Acknowledgements

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

The authors gratefully acknowledge the US National Data Center for providing the NVAR array data and thank William Walter, Stephen Myers, Stanley Ruppert, Tormod Kvaerna, Steven Gibbons, William Junek, for helpful discussions, encouragement and support. The authors also thank Teresa Hauk for help with data selection and collection.

References

Gibbons, S. J. and F. Ringdal (2006), The detection of low-magnitude seismic events using array-based waveform correlation, *Geophysical Journal International*, **165**, 149-166.

Gibbons, S. J., M. B. Sorensen, D. B. Harris and F. Ringdal (2007), The detection and location of low magnitude earthquakes in northern Norway using multi-channel waveform correlation at regional distances, *Phys. Earth Planet. Inter.*, **160**, 285-309.

Harris, D. (2006), Subspace detectors: Theory, Lawrence Livermore Natl. Lab. Rep. UCRL-TR-222758, 46 pp., Lawrence Livermore Natl. Lab., Livermore, Calif. (Available at http://www.llnl.gov/tid/lof/documents/pdf/335299.pdf) [proposed for electronic supplement]

Harris, D., and T. Paik (2006), Subspace detectors: Efficient implementation, Lawrence Livermore Natl. Lab. Rep. UCRL-TR-223177, 36 pp., Lawrence Livermore Natl. Lab., Livermore, Calif. (Available at http://www.llnl.gov/tid/lof/documents/pdf/336400.pdf) [proposed for electronic supplement]

Hauksson, E., D. Oppenheimer, T. M. Brocher (2004). Imaging the source region of the 2003 San Simeon earthquake within the weak Franciscan subduction complex, central California, *Geophys. Res. Lett.* **31**, L20607, doi 10.1029/2004GL021049

Israelsson, H. (1990), Correlation of waveforms from closely spaced regional events, *Bull. Seismol. Soc. Am.* **80**(6), 2177-2193.

Lay, T. and T. Wallace (1995), *Modern Global Seismology*, Academic Press (ISBN 0-12-732870-X), San Diego, California.

Maceira, M., C. A. Rowe, G. Beroza, and D. Anderson (2010), Identification of low-frequency earthquakes in non-volcanic tremor using the subspace detector method, *Geophys. Res. Lett*., 37, L06303.

McLaren, M. K., J. L. Hardebeck, N. van der Elst, J. R. Unruh, G. W. Bawden and J. L. Blair (2008). Complex faulting associated with the 22 December 2003 Mw 6.5 San Simeon, California, earthquake, aftershocks, and postseismic surface deformation, *Bull. Seismol. Soc. Am.* **98**(4), 1659-1680.

Nadeau, R. M., W. Foxall, and T. V. McEvilly (1995), Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, California, *Science*, **267**, 503-507.

Nadeau, R. M., and T. V. McEvilly (1997), Seismological studies at Parkfield V: characteristic microearthquake sequences as fault- zone drilling targets, *Bull. Seismol. Soc. Am.*, **87**(6), 1463-1472.

Schaff, D. P. (2008) Semiempirical statistics of correlation-detector performance, *Bull. Seismol. Soc. Am.*, **98**(3), 1495-1507

Schaff, D. P. (2009), Broad-scale applicability of correlation detectors to China seismicity, *Geophys. Res. Lett.*, 36, L11301.

Schaff, D. P. and P. G. Richards (2004), Repeating seismic events in China, *Science*, **303**, 1176-1178.

Schaff, D. P., and F. Waldhauser (2005), Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, Bull. Seismol. Soc. Am., **95**, 2446– 2461.

Scharf, L. L. and B. Friedlander (1994), Matched subspace detectors, *IEEE Trans. on Signal Processing*, **42**(8), pp. 2146-2157.

VanDecar, J. C. and R. S. Ccrosson (1990), Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seismol. Soc. Am.,* **80**(1), pp 150-169.

Wessel, P. and W. H. F. Smith (1991), "Free software helps map and display data," *EOS Trans. AGU*, **72**, 441.

Figure Captions

Figure 1 An automated detection framework was developed to spawn correlation (subspace) detectors to group events in a developing aftershock sequence. This simplified block diagram includes a detector list, with fixed beamformers (one only used in this study) and dynamically added subspace detectors. The system has a database of detected events which is used to consolidate subspace detectors at periodic intervals.

Figure2 The main sequence of study consists of the magnitude 6.5 San Simeon earthquake and its aftershocks during the 10 day period Dec 22, 2003 – Dec 31, 2003. The system of Figure 1 was used to process 10 days of data from the NVAR array. The array is approximately 390 kilometers from the main shock, with the geometry shown in the inset.

Figure3 Signals from the San Simeon aftershocks had durations of about 100 second. A long (110 second) window indicated by the red bars was used to form subspace detector templates. The green bar indicates the STA/LTA detection time corrected for an estimated lag in the short-term average.

Figure 4 Superimposed events can produce conflicting subspace detectors. This particular group of two events (A and B) produced two templates originating at the P onsets of A and B. Because the signal from B is so much larger, it dominated both templates and gave rise to large numbers of duplicate detections. Safeguards were added to the system to prevent this behavior.

Figure 5 Regional events unrelated to the San Simeon sequence were used to bound the detection threshold for subspace detectors. This map shows locations of the 14 regional events

used to compute a sample correlation distribution. The San Simeon main shock is indicated by the red star.

Figure 6 Waveforms recorded by station NV08 of the NVAR array for the 14 events used in the correlation calculation, show that the durations and S-P times of the signals are approximately the same for all events. The red box indicates the portions of the waveforms used in the correlation calculation, which have the same durations as detector templates.

Figure 7 The subspace (correlation) threshold was set substantially higher than the distribution of (squared) correlation values for the ensemble of unrelated regional event . The correlations were calculated over 9 channels of NVAR data.

Figure 8 The detection framework detected substantially more San Simeon events than a beamformer directed at the sequence (Pn velocity and great circle back azimuth), but organized them into a smaller number of groups (184 vs 306). The ANSS locations of events in the source region are shown in grey. Events detected by both the beamformer and the spawned subspace detectors are shown in black. Events detected just by the subspace detectors are shown in red.

Figure 9. Event histograms by magnitude show that the detection framework mostly added events below magnitude 3, though some events were added above magnitude 3. The histogram of ANSS catalog events is shown in grey. That of the beamformer is shown in black and the detection framework in red.

Figure 10 San Simeon aftershocks associated with one particular subspace detector (#116) show that all associated waveforms have broad similarities, though there are differences in

detail. The variations are caused by the relatively low (0.316) correlation threshold. The events, nonetheless, are related.

Figure 11 San Simeon aftershocks associated with two additional detectors (#110 and #111) demonstrate the same pattern of waveform similarity within a group, but substantially greater variation between groups.

Figure 12 Events within automatically-defined groups tend to form compact geographic clusters. This fact supports the notion that clustering by waveform correlation assembles related events that may be interpreted efficiently as a group. Group membership is indicated by color and referenced to particular detectors by number.

Figure 13 The detection framework also groups local events automatically, which could aid efficient interpretation during normal operations. Based on what appear to be variable infrasound arrivals lat in the record, these events may be mining explosions.

Figure 14 Waveforms with a small time-bandwidth product associated with a local source (or simply noise bursts) may be relatively simple to dismiss when organized as a group. This is the largest such group produced by the detection framework.

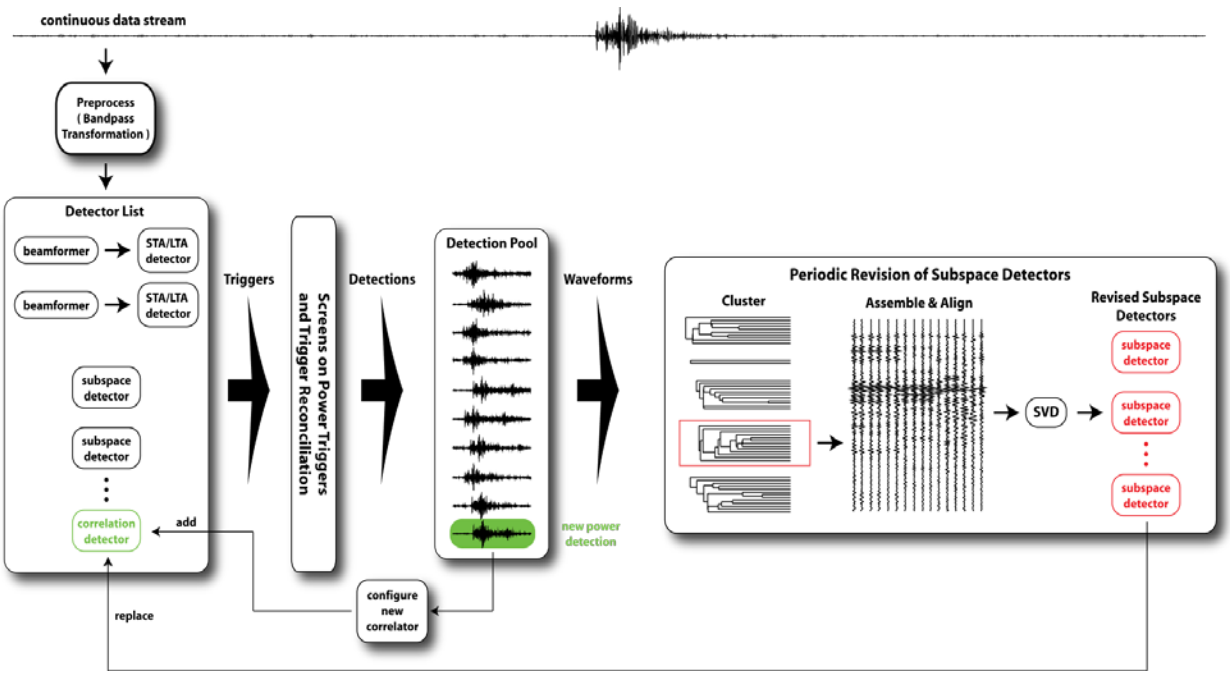
continuous data stream
Preprocess (Bandpass Transformation)
Detector List
beamformer
STA/LTA detector
beamformer
STA/LTA detector
subspace detector
subspace detector
correlation detector
add
Triggers
Screens on Power Triggers and Trigger Reconciliation
Detections
Detection Pool
Waveforms
new power detection
configure new correlator
replace
Periodic Revision of Subspace Detectors
Cluster
Assemble & Align
SVD
Revised Subspace Detectors
subspace detector
subspace detector
subspace detector

Figure 1

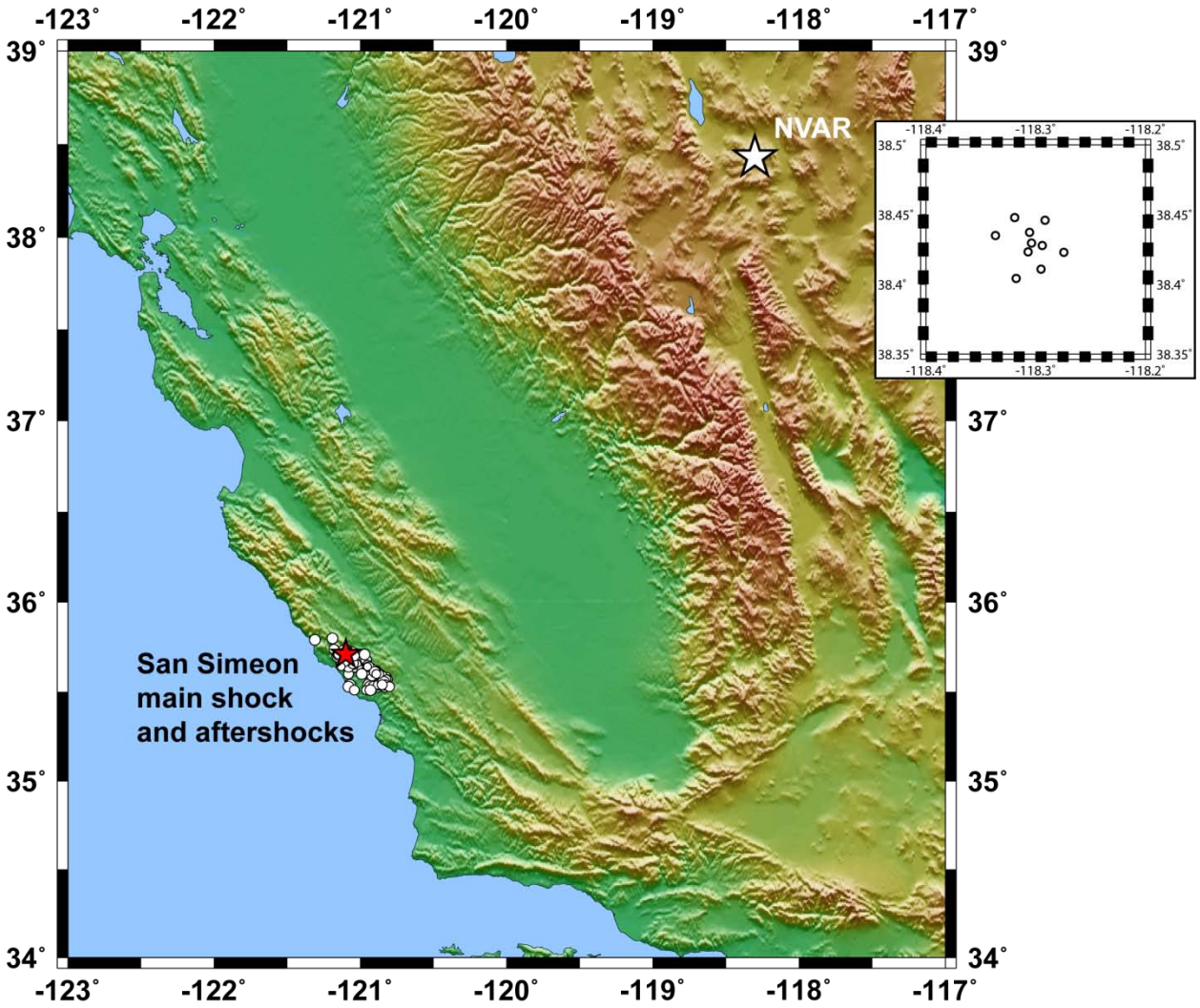
-123˚
-122˚
-121˚
-120˚
-119˚
-118˚
-117˚
39˚
38˚
37˚
36˚
35˚
34˚
NVAR
San Simeon
main shock
and aftershocks
-118.4˚
-118.3˚
-118.2˚
38.5˚
38.45˚
38.4˚
38.35˚

Figure 2

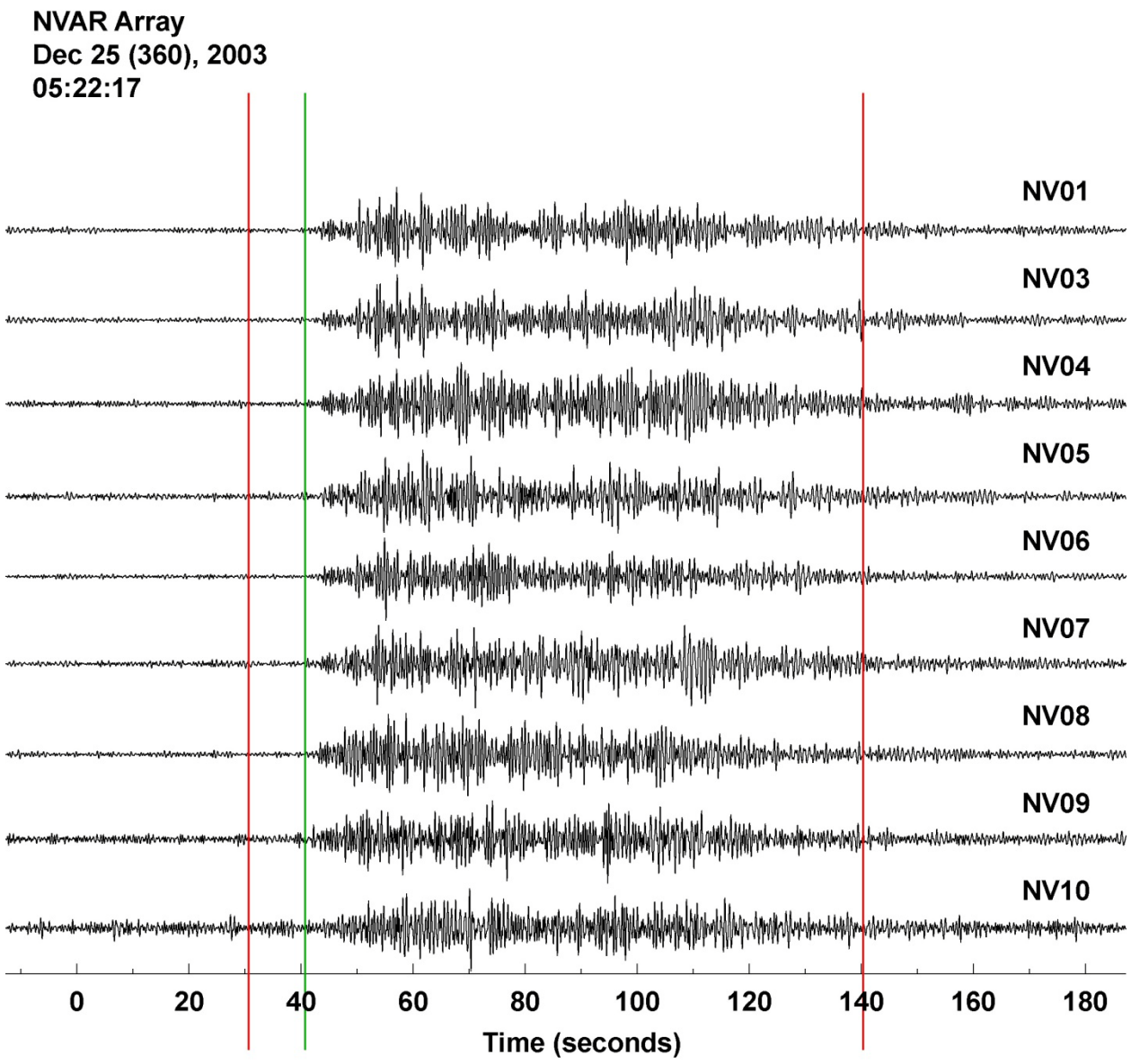
NVAR Array
Dec 25 (360), 2003
05:22:17
NV01
NV03
NV04
NV05
NV06
NV07
NV08
NV09
NV10
0
20
40
60
80
100
120
140
160
180
Time (seconds)

Figure 3

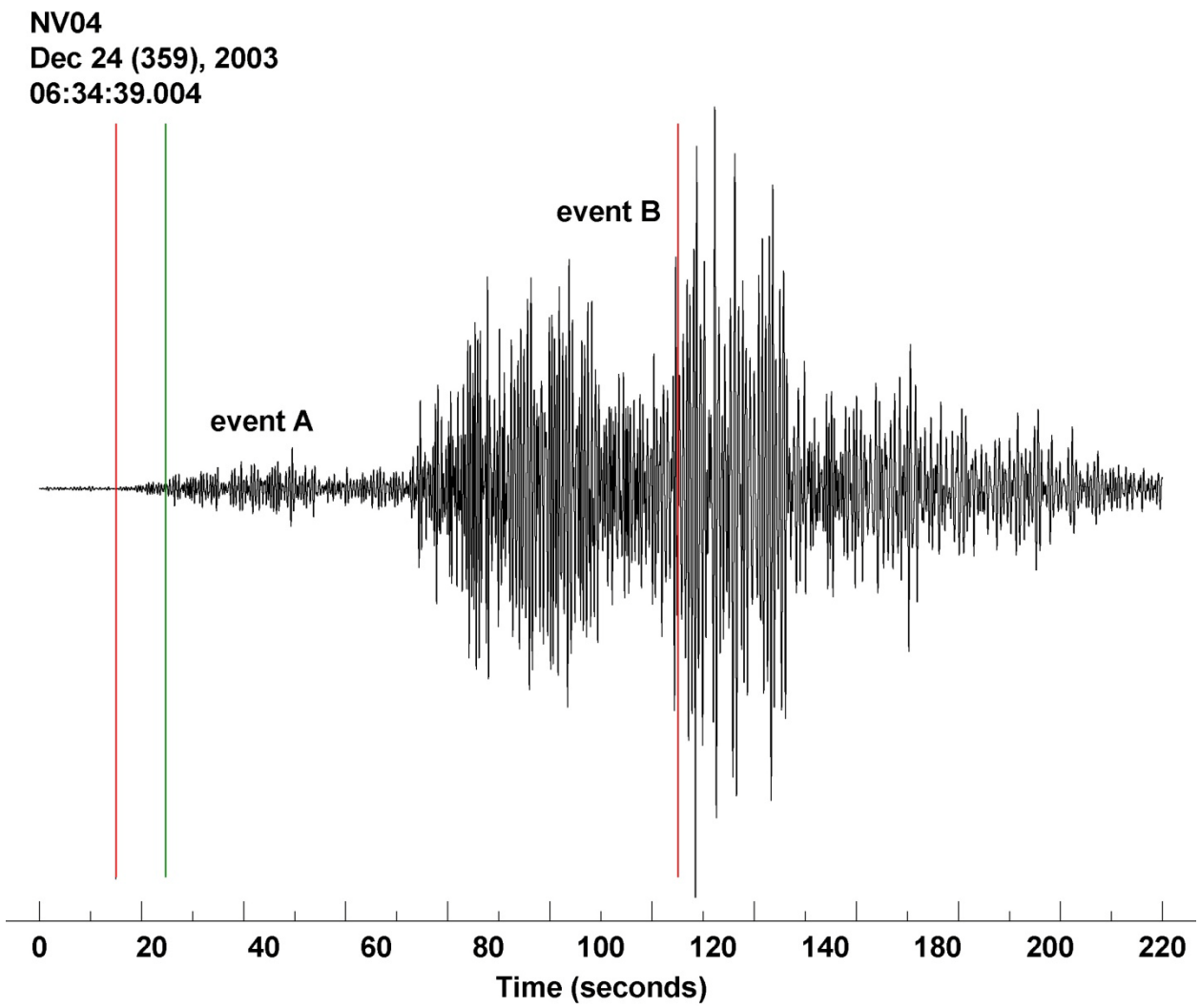
NV04
Dec 24 (359), 2003
06:34:39.004
event B
event A
0
20
40
60
80
100
120
140
180
200
220
Time (seconds)

Figure 4

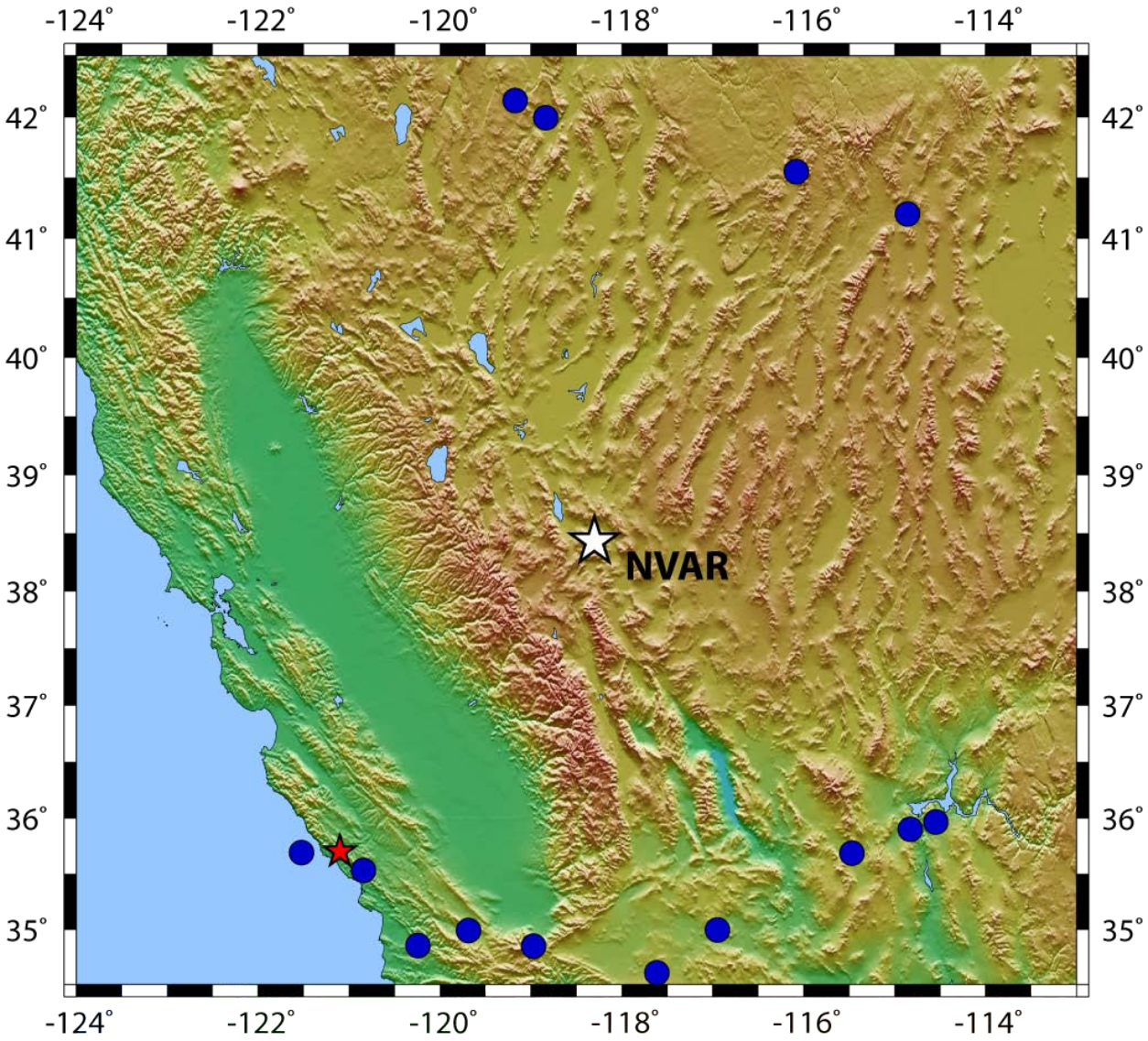

Figure 5

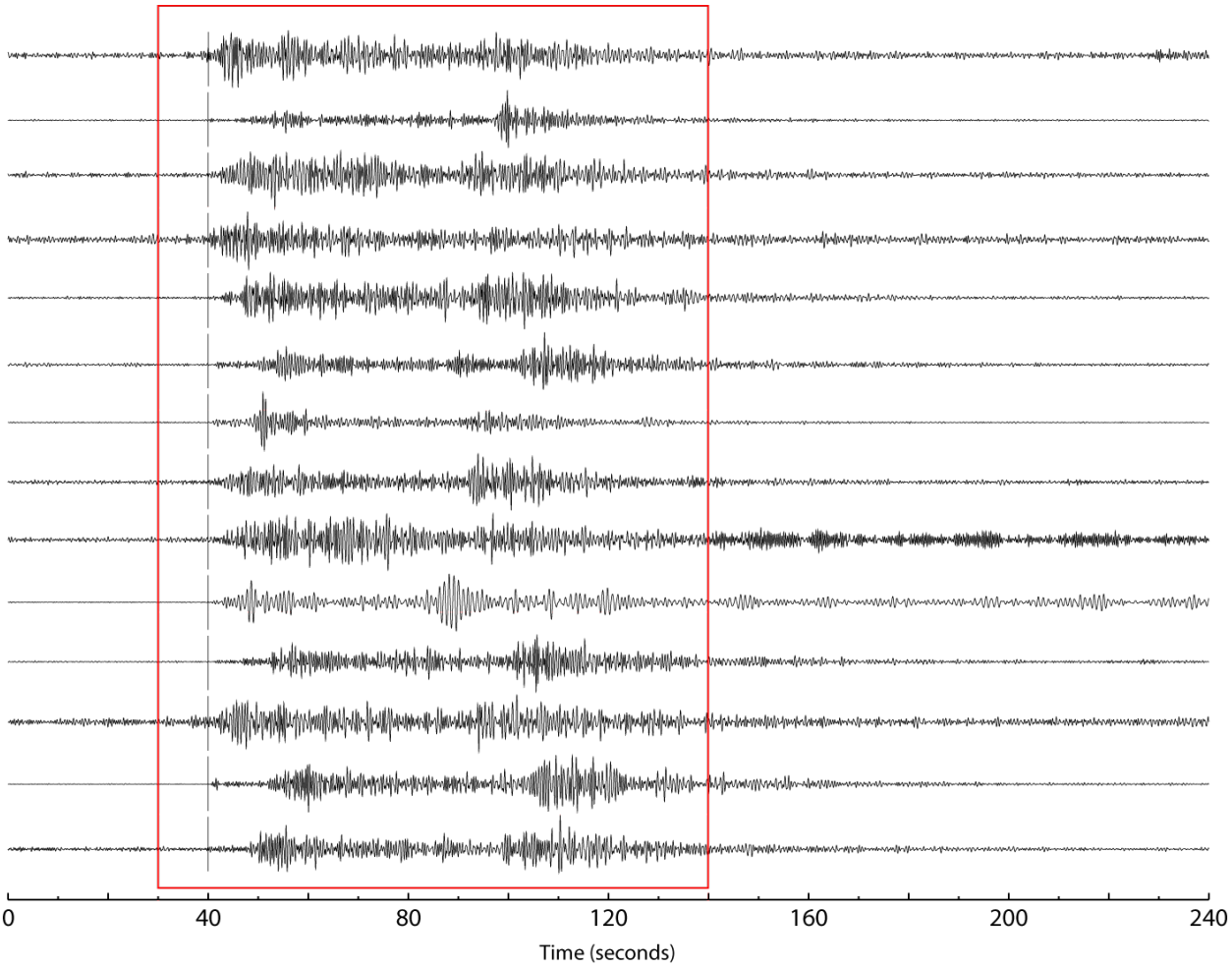

Figure 6

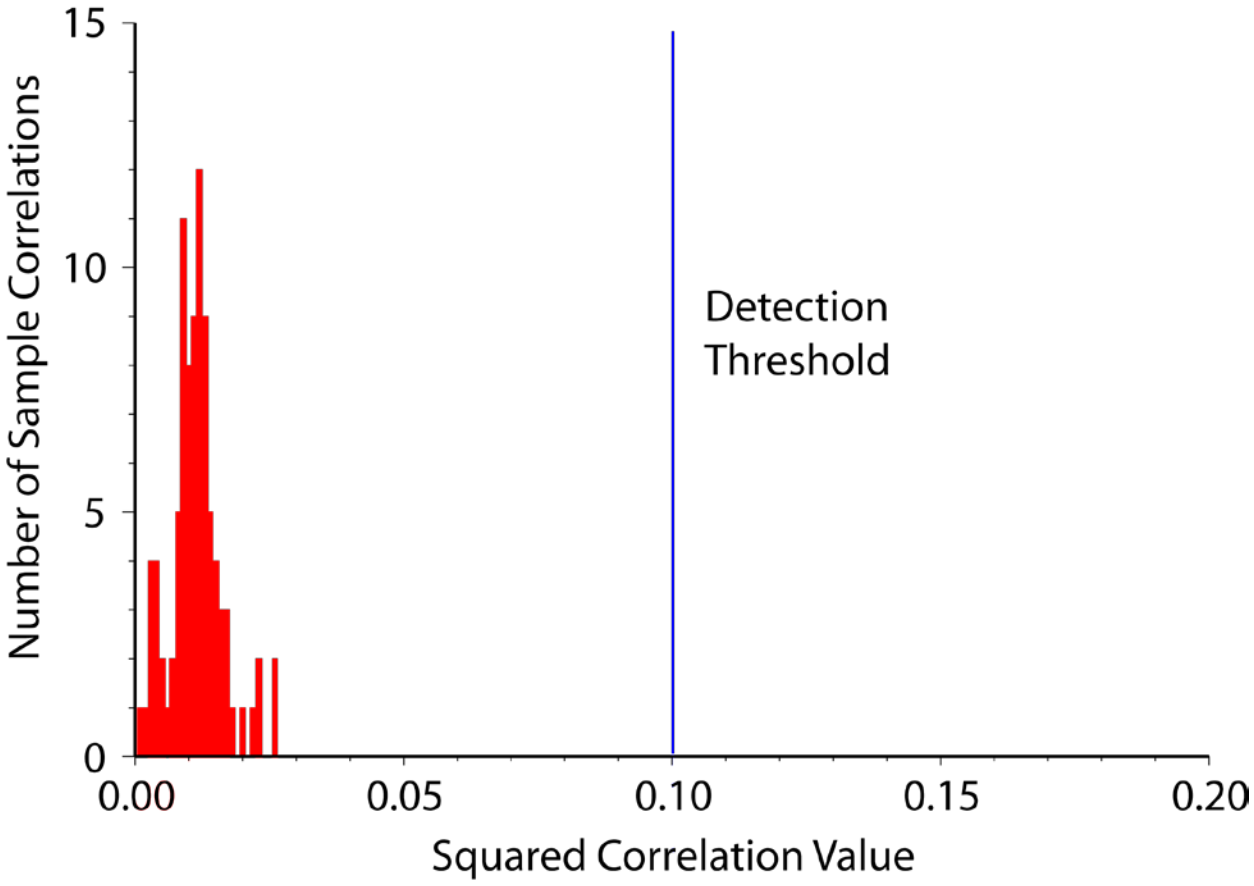
15
10
5
0
Number of Sample Correlations
Detection
Threshold
0.00
0.05
0.10
0.15
0.20
Squared Correlation Value

Figure 7

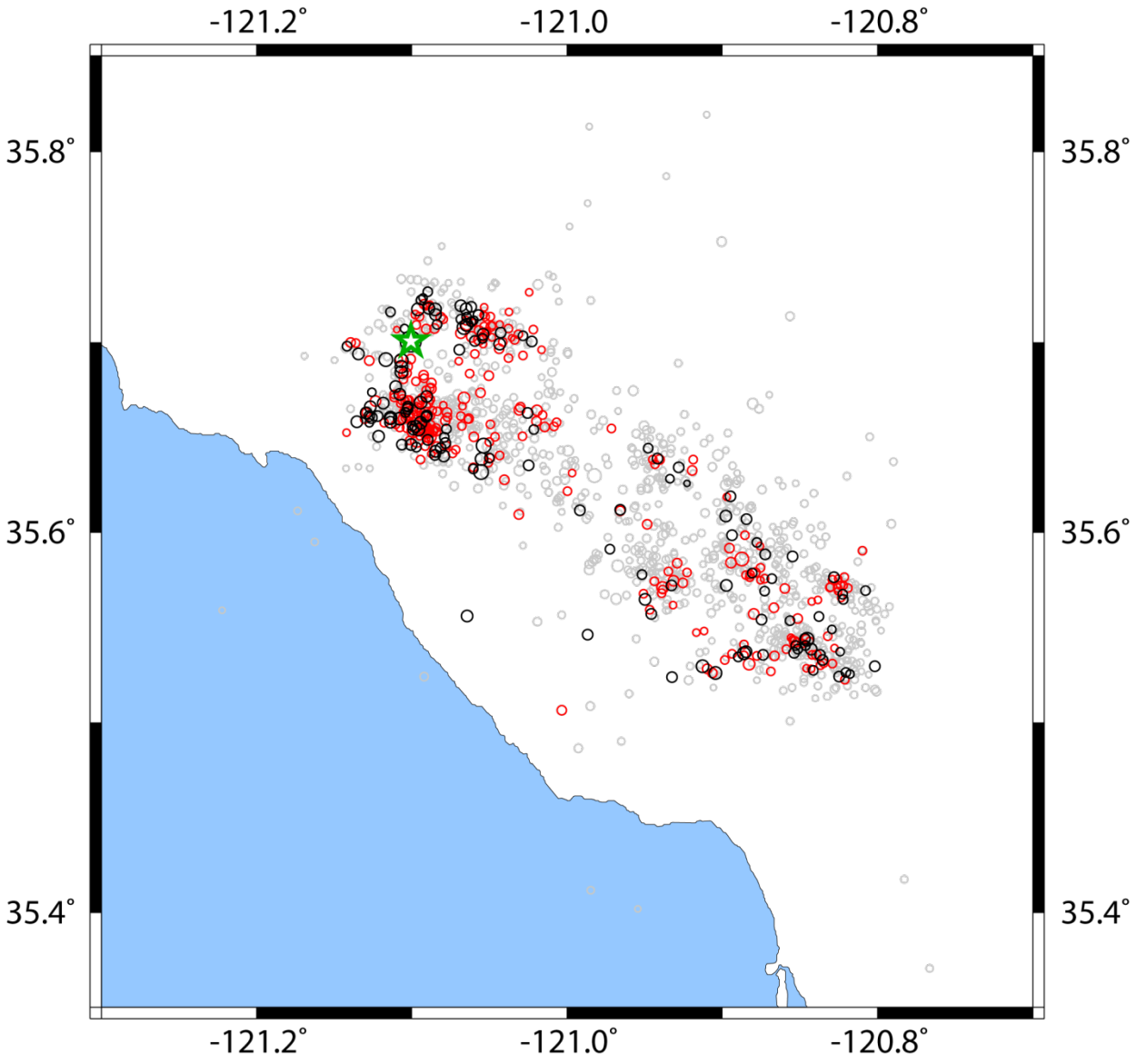

Figure 8

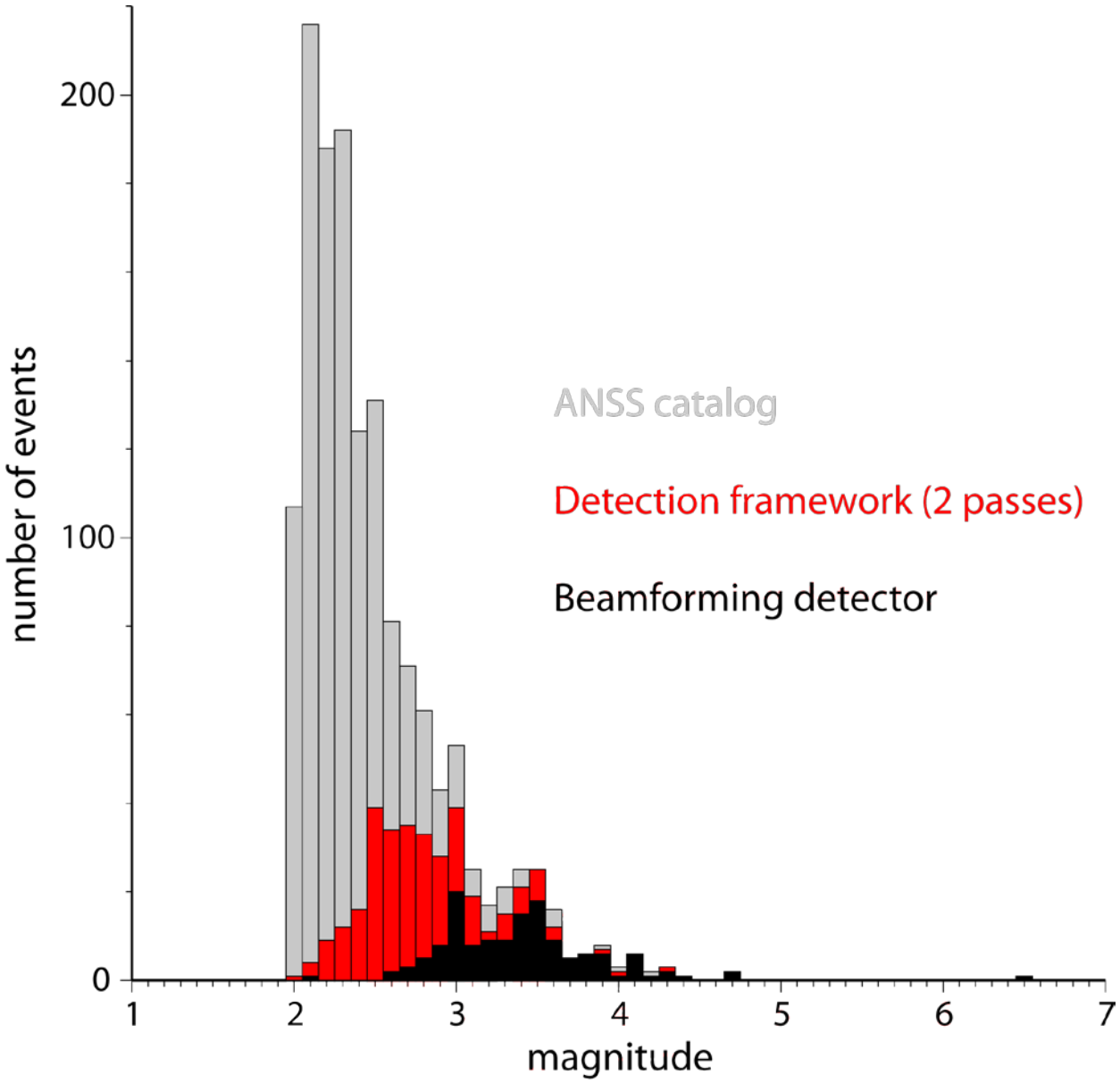
200
100
0
number of events
1
2
3
4
5
6
7
magnitude
ANSS catalog
Detection framework (2 passes)
Beamforming detector

Figure 9

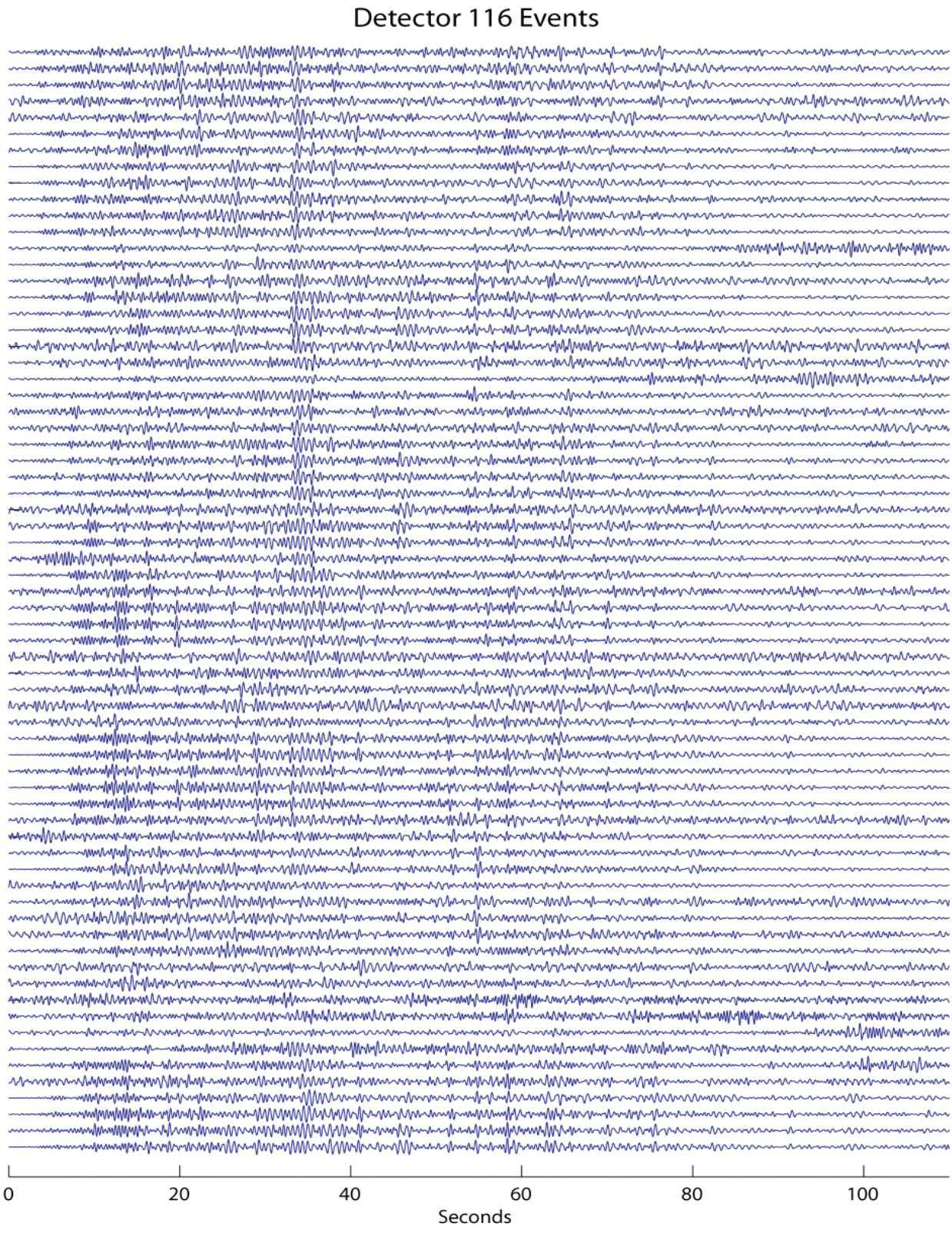

Figure 10

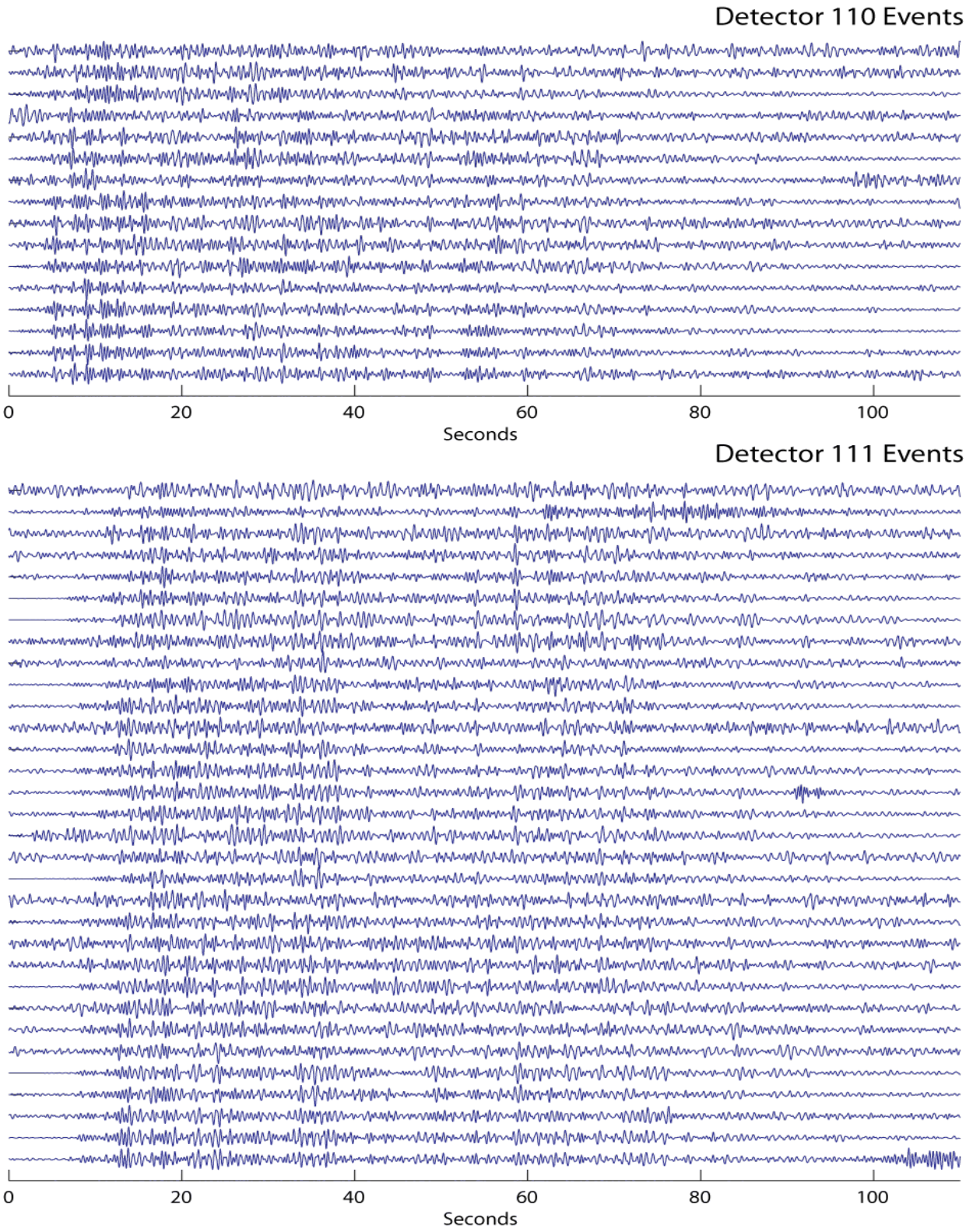
Detector 110 Events
0
20
40
60
80
100
Seconds
Detector 111 Events
0
20
40
60
80
100
Seconds

Figure 11

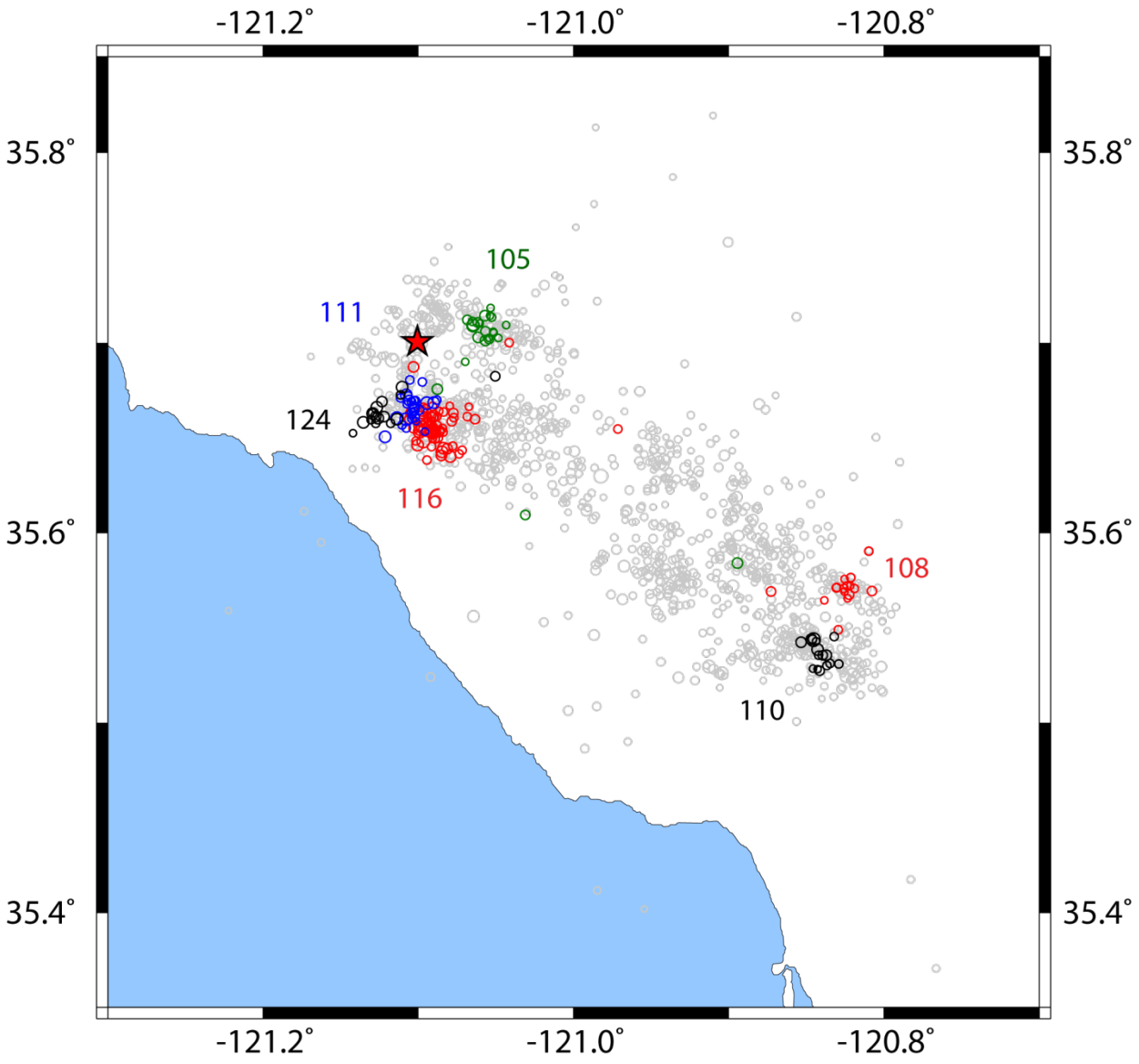

Figure 12

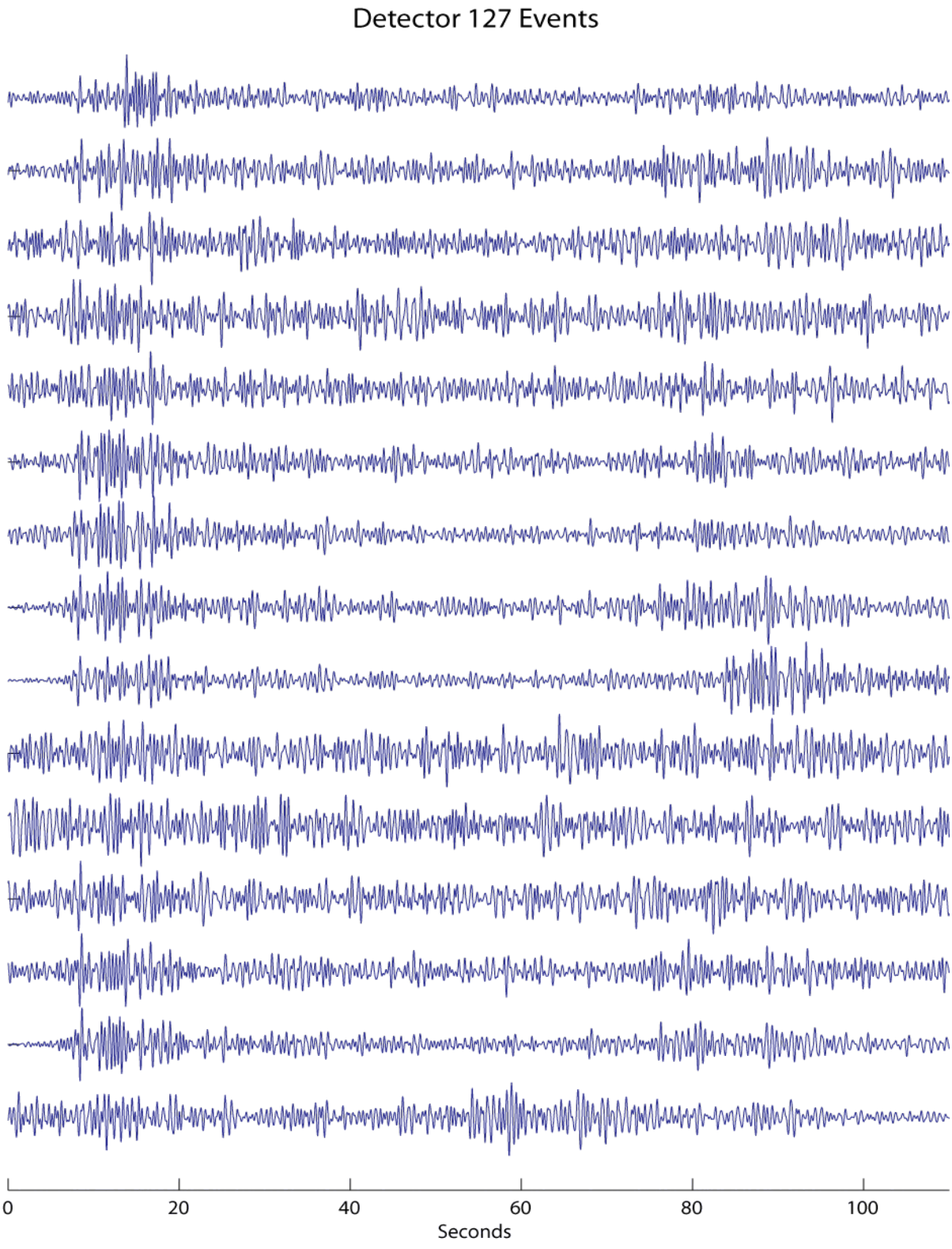

Figure 13

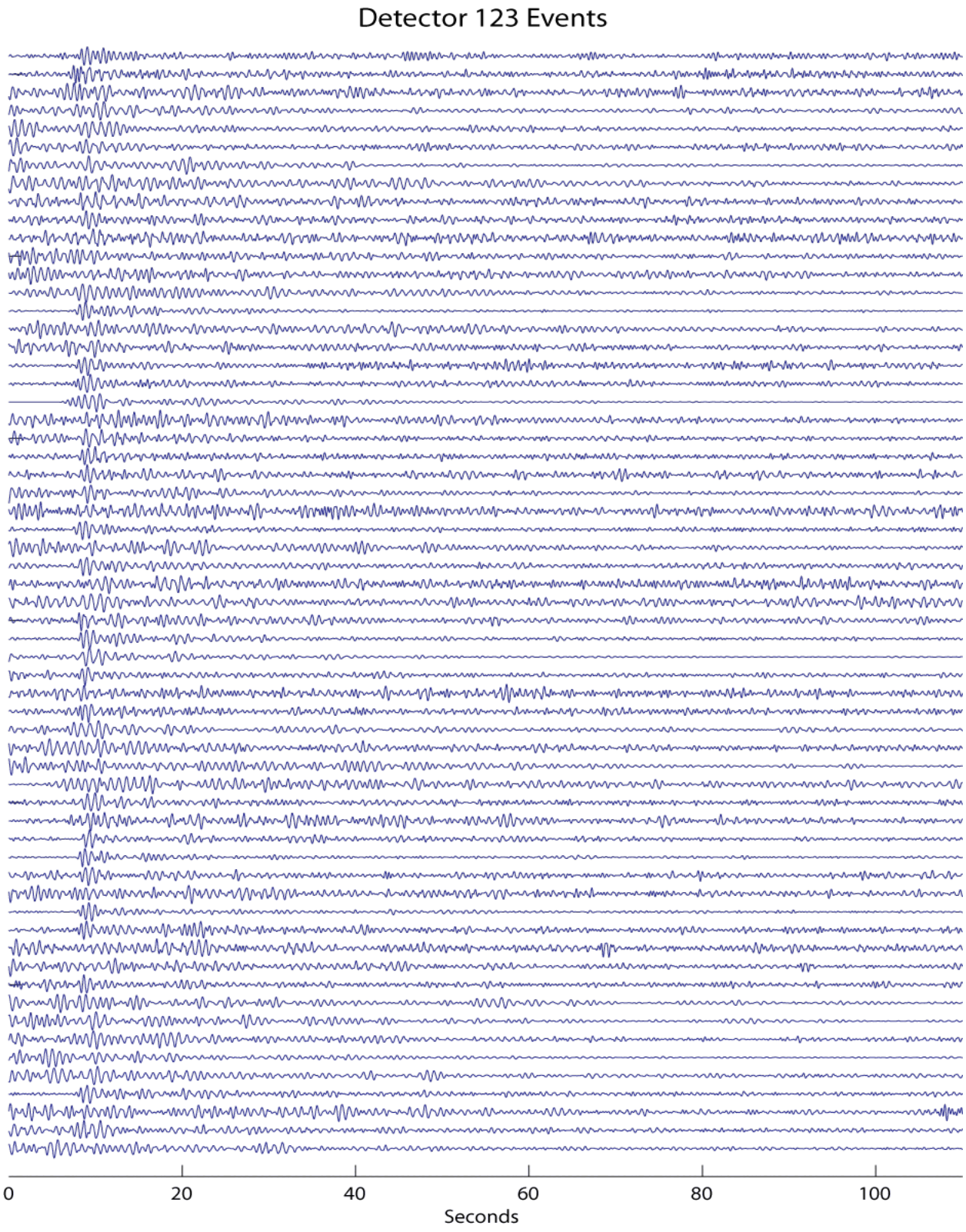

Figure 14